

Our Ref.: **871-81**
TT/15/US

U.S. PATENT APPLICATION

Inventor(s): Guy Nathan
Dominique Dion

Invention: PROCESS FOR ADJUSTING THE SOUND VOLUME OF A DIGITAL
SOUND RECORDING

NIXON & VANDERHYE P.C.
ATTORNEYS AT LAW
1100 NORTH GLEBE ROAD
8TH FLOOR
ARLINGTON, VIRGINIA 22201-4714
(703) 816-4000
Facsimile (703) 816-4100

SPECIFICATION

PROCESS FOR ADJUSTING THE SOUND VOLUME OF A DIGITAL SOUND
RECORDING

Field of the invention

The present invention relates to a process for adjusting the sound volume of a digital sound recording reproduced by an item of equipment. This process is essentially intended to be used during the reproduction of a digital recording in the form of a data file by means of a sound card, for example, of an audiovisual reproduction system, such as a jukebox.

Background of the invention

In the prior art, it is known that digital recordings, such as compact disks (CD), are not reproduced with the same sound volume for a specified sound setting level. This is essentially due to the type of music and the way in which the piece of music was recorded. Indeed, a sound frame is composed of an electrical signal comprising a succession of oscillations and peaks. Each peak corresponds to a voltage value. The higher the voltage in terms of absolute value, the higher the volume and the higher the slope of the signal variation, the higher the frequency of the sound reproduced. When such a recording is recorded in the form of a digital file and then reproduced on a sound system by means of a digital sound card on a computer, the same maximum variation phenomena are observed since the data contained in the file is approximately the same as that recorded on a CD. Consequently, between two recordings of different types of music, it is necessary to modify the

sound level setting between two recordings, to obtain a reproduction with the same sound level for two different recordings with different original sound levels.

5 Objects and summary of the invention

Therefore, the purpose of the present invention is to remedy the disadvantages of the prior art by proposing a process for adjusting the sound level of a digital sound recording making it possible to obtain identical
10 sound levels in different recordings, irrespective of the differences in the digital sound recording level existing initially between each of the recordings.

This purpose is achieved by the fact that the process comprises:

15 - a step consisting of determining, in absolute values, for a recording, the maximum amplitude values for sound frequencies audible for the human ear,

 - a step consisting of calculating the possible gain for a specified sound level setting, between the maximum
20 amplitude value determined above and the maximum amplitude value for all frequencies combined,

 - a step consisting of reproducing the recording with a sound card by automatically adjusting the amplification gain level making it possible to obtain a
25 sound level for the recording of a specified value so that it corresponds to the gain calculated for this recording.

According to another feature, the maximum amplitude value determination step comprises:

30 - a step consisting of counting the number of samples of the recording with a specified amplitude, for all the amplitudes existing in the recording,

- a step consisting of classifying the amplitudes of the number of samples found in increasing order,

- a step consisting of storing in memory the maximum amplitude, for all frequencies combined, and the amplitude, for which the order number in the classification carried out is n ranks less with reference to the rank of the maximum amplitude, the amplitude found corresponding in this case to the maximum amplitude for frequencies audible for the human ear.

10 According to another feature, n is determined so that the degradation of the reproduction quality of the recording is not perceptible to the human ear.

According to another feature, n is of the order of 10 and preferably equal to 4 or 5.

15 According to another feature, the maximum amplitude value determination step comprises:

- a step consisting of counting the number of samples of the recording with a specified amplitude, for all the amplitudes existing in the recording,

20 - a step consisting of classifying the amplitudes of the number of samples found in increasing order,

- a step consisting of calculating the mean value A_{mean} of the n' highest amplitudes occurring at least k' times in the recording.

25 According to another feature, the maximum amplitude value determination step comprises:

- a step consisting of compressing the recording by means of at least one psycho-acoustic mask making it possible to eliminate inaudible sounds from the initial

30 recording,

- a step consisting of decompressing the recording,

- a step consisting of searching the maximum amplitude on the decompressed recording, this amplitude corresponding in this case to the maximum amplitude for frequencies audible for the human ear.

5 According to another feature, the psycho-acoustic mask(s) is /are applied using the MPEG-1 Layer 3 process.

According to another feature, the reproduction step comprises a dynamic reproduction sound level adjustment step on the recording consisting of authorising a
10 specified gain for the low-pitched and/or high-pitched sounds in the recording, the gain corresponding approximately to the attenuation applied during the reproduction of the recording.

Another purpose of the invention consists of a use
15 of the process according to the invention in an audiovisual reproduction system, such as a digital jukebox.

This purpose is achieved by the fact that the automatic volume adjustment process is used on a digital
20 audiovisual reproduction system, this use being characterised in that the recording is stored in memory in the reproduction system with the corresponding calculated gain and audiovisual reproduction system reading means giving access to the gain value to control
25 the gain circuits of the digital signal processing processor of the digital audiovisual reproduction system to adjust the sound level accordingly.

Brief description of drawings

30 Other features and advantages of the present invention will be understood more clearly upon reading

the description below with reference to the appended drawings, wherein:

- Figure 1 represents a block diagram of a sound card using the process according to the invention,

5 - Figures 2A and 2B represent a curve representing the frequency of the occurrence of a voltage in a digital recording,

- Figure 3 represents a sound frame of a recording.

10 Description of the preferred embodiments

Before starting the description of the invention, it is necessary to give some notes on digital recording. First of all, sound reproduction by a loud speaker consists of applying voltages of specified levels to said
15 loud speaker, according to a specified frequency to vibrate a membrane and, therefore, produce the sound corresponding to the specified frequency. For a given amplification value, the root mean square voltage value defines the sound volume or sound level.

20 A sound frame, represented in Figure 3, is therefore formed by superimposing oscillations representing the variations over time of the amplitude of the power supply voltage of an acoustic reproduction component such as a loud speaker. The digitisation of a sound recording
25 consists, in fact, of performing sampling of the sound frame and, therefore, reading the voltage values according to time intervals determined by a periodicity. The shorter the period, the more precise the digitisation. During the reproduction of the recording,
30 the analogue signal is reconstructed from digital samples stored during the digitisation. The dots on the curve represent the samples used during the digitisation.

In this way, depending on the type of music, the curve C representing the frequency of the reproduced sound, defined by the slope of the curve C and the corresponding voltage value of the maximum sound levels, for the same sound amplification circuit setting, the output level of the loud speakers will be different. Indeed, the maximum root mean square voltages observed for a first recording will not necessarily be of the same order as the maximum root mean square voltages observed for a second recording. Therefore, the purpose of the invention is to provide a solution for this disadvantage such that, between two recordings, the volume or sound level perceived by the listener is automatically adjusted so that the sound level is the same from one recording to another.

The invention requires, firstly, a preliminary analysis of each recording liable to be reproduced on an audiovisual reproduction system or on a computer and, secondly, a correction of the amplification level during the sound reproduction of the recording, according to the analysis.

A first solution consists of searching, in absolute values, the maximum voltage observed on each recording, and using this value to amplify the recordings such that, for a specified sound level setting, this values reaches the same voltage value for all the recordings. However, a sound frame of a recording comprises sounds with frequencies that are both audible and inaudible for the human ear. In this way, if the maximum amplitude corresponds to an inaudible frequency, the adjustment of the volume will not be adapted.

Therefore, the process according to the invention consists, in a first step, of determining, for a recording, the maximum amplitude only for frequencies audible for the human ear. In a first embodiment variant, 5 this maximum amplitude is determined by analysing the digital recording to classify the number of samples of the recording for each amplitude, in increasing order of amplitude, in absolute values. This classification is represented in Figure 2A. The Y-axis represents the 10 number N of occurrences of a specified amplitude in the recording and the X-axis represents in absolute values the number representing the voltage of the analogue signal in volts during the digital encoding of the analogue signal with a precision of 16 bits. During the 15 digitisation of the analogue signal representing a sound signal, each sampled voltage is encoded with a number between -32767 and 32767 when the precision is 16 bits.

Empirically, it is observed that a recording corresponding to a song only comprises a few samples, of 20 the order of ten, located in the portion B of the curve C1, with the highest amplitudes in the recording. In this way, the portion B of the curve C1 is represented with dashes to show that all the values of the numbers representing the voltages of the corresponding analogue 25 signal are not represented. Similarly, it is observed that 90% of the samples of a recording have a low amplitude, i.e. located in the portion A of the curve C1.

According to the invention, the maximum amplitude is selected, in the classification carried out, as the 30 amplitude n ranks less with reference to the rank of the maximum amplitude sample of the recording. In other words, if 1 corresponds to the rank of the number

representing the amplitude and K is the rank of the number representing the maximum amplitude found on the digital recording, then the amplitude selected as the maximum amplitude for the process corresponds to the rank number K-n, from the classification defined and corresponding to the curve C1. In this way, the n-1 samples, located on portion B of the curve C1 are not taken into account, using the maximum amplitude as a basis, implying that these samples do not appear in the final reproduction. Then, the recording volume correction, i.e. the possible volume gain Gv for the recording is determined by applying the following formula:

$$Gv = 20\log(A_2/A_m) \quad a$$

In this formula, A_2 is the selected amplitude and A_m is the maximum amplitude of the recording.

In practice, the higher the value of n, the more degraded the recording reproduction quality. Indeed, the higher the value of n, the higher the number of high-amplitude samples that will not be taken into account, and the higher the probability of the samples not taken into account corresponding to audible signals. Consequently, when the gain calculated using the above formula is applied to the recording, some sound frequencies will be over-amplified, resulting in a saturation phenomenon on the loud speakers and, therefore, in a degradation of the reproduction quality. It has been observed that a value of n of the order of 10, preferably equal to 4 to 5, does not induce a perceptible degradation during the reproduction of the recording after applying the gain calculated using the formula above. This variant can only be applied

effectively to digital recordings that have not undergone prior compression or processing aiming to optimise the volume level.

On the basis of the classification carried out above, another variant for determining the value of the selected amplitude may be carried out. According to this variant, the value of the selected amplitude corresponds to the mean value A_{mean} of the n' highest amplitudes occurring at least k' times in the recording. Then, the value of the possible volume gain G_v for the recording is determined by applying the formula a above, replacing A_2 by A_{mean} .

The experiment showed that, by choosing n' equal to 2 and k' equal to 4, the sound recording reproduction did not show any degradation audible for the human ear. The higher the values of n' and k' , the higher the degradation of the sound recording reproduction.

Figure 2B represents the result of the classification carried out on a recording having undergone processing aiming to optimise the sound level. Indeed, recordings that have undergone this type of processing already take the presence of inaudible frequencies into consideration and tend to eliminate these frequencies for the benefit of improved recording volume management. For these specific recordings having undergone processing to optimise the volume, since the number of samples with a high amplitude value is higher, these amplitudes also correspond to audible signals. Consequently, the step described above is applicable but results in a perceptible degradation of the reproduction of the recording.

For recordings having undergone optimisation processing, the determination step of the maximum amplitude for audible signals consists of compressing the recording according to a compression process using at least one psycho-acoustic mask making it possible to eliminate inaudible sounds from the recording. For example, it is possible to use the known MPEG-1 Layer 3 process or any other compression process such as AAC. Indeed, it is known that the MPEG compression process uses masks to eliminate any unnecessary data from the recording. The unnecessary data in the sound recording includes all the inaudible frequencies and all the sound variations which are not perceptible to the human ear. Then, the recording is decompressed and the value of the maximum amplitude is located in this decompressed recording. In this way, during the decompression, the decompressed recording only contains audible frequency sounds. Searching the maximum amplitude in this decompressed recording does not necessarily produce a maximum amplitude A_m for an audible frequency. In this embodiment variant, it is also advisable to store in memory before compression, the maximum amplitude of the recording, for all frequencies combined, in order to be able to calculate the gain according to the formula a. This second embodiment variant may be applied to any type of recording, since the MPEG compression process is indifferent to the initial recording type.

The gain value calculated by means of the formula a is then stored in memory with the sound recording produced, for example, on a server or on the audiovisual reproduction system, and used during the recording reproduction by the reproduction system. Indeed, during

the reproduction of the initial digital recording, the gain calculated for this recording is added during the sound setting.

The process according to the invention is particularly used when digital recordings are reproduced by means of a sound card of a computer or an audiovisual data reproduction system. Therefore, the process according to the invention requires having determined the gain either arbitrarily or using a preliminary analysis of each recording liable to be reproduced by the sound card. As described above, this analysis consists of determining the gain liable to be applied to each recording during its reproduction. The gain is, for example, stored in memory in a database on storage means of the computer or reproduction system and can be accessed by the sound card management program, such that each recording stored on the storage means of the computer or the reproduction system is associated with a gain in the database. In this way, before the reproduction of a specified recording, the sound card management program consults its database and collects the data representing the gain calculated for this recording. During the setting of the sound of the recording, the level selected by the user will be automatically adjusted by a value corresponding to the calculated gain G_v , such that the real sound level indeed corresponds to the level selected by the user and is homogeneous for all the recordings contained in the storage means. The adjustment may be made by a positive or negative value.

Another variant of the process according to the invention consists of adjusting the gain for the sound signals of a recording corresponding to low-pitched

and/or high-pitched sounds. The aim of the process is to increase, when possible, the gain for low-pitched and/or high-pitched sounds without exceeding the sound level selected by the user and without exceeding a maximum gain set for low-pitched and/or high-pitched sounds. It is necessary to underline that, in this variant, only low-pitched and/or high-pitched sounds are concerned by the dynamic gain adjustment, when the reproduction enables independent setting of the general sound level and the sound level of low-pitched and/or high-pitched sounds. In this way, when the sound level of low-pitched and/or high-pitched sounds is less than the sound level selected by the user, an additional gain is authorised on low-pitched and/or high-pitched sounds to increase the perception of low-pitched and/or high-pitched sounds to improve the reproduction quality of the recording. This additional gain will be at most equal to the gain requested by the user for low-pitched and/or high-pitched sounds.

The maximum volume is obtained when the incoming signal on the amplifier is not attenuated, i.e. at a gain of 0 dB. So as to obtain a gain for low-pitched and/or high-pitched sounds systematically, the overall maximum volume for the recording may be less than zero dB and the maximum volume of low-pitched and/or high-pitched sounds is determined so that the incoming gain in the amplifier can be equal to zero dB. Consequently, it is always possible to obtain a gain for low-pitched and/or high-pitched sounds corresponding to the absolute value of the recording volume attenuation. In this way, for example, if the recording volume attenuation is -3 dB, the gain for low-pitched and/or high-pitched sounds is 3 dB. So as

to limit the influence of the dynamic adjustment of low-pitched and/or high-pitched sounds, the maximum low-pitched and/or high-pitched sound gain is limited, for example to 12 dB. In this way, even if, for a specified
5 volume, the gain for low-pitched and/or high-pitched sounds may be 16 dB, for example, it will only actually be 12 dB.

For example, Figure 1 represents a block diagram of a sound card using the process according to the
10 invention. This sound card is connected, for example, to a central processing unit (not shown) of a computer or a reproduction system comprising, particularly, storage means in which a sound card management program, or pilot, is particularly stored. The sound card represented in
15 Figure 1 comprises, for example, 3 inputs 11, 12, 13. A first input 11 receives the signals representing the recordings, for example, through an MPEG decoder, the second input 12 receives signals from an auxiliary source and the third input 13 receives signals from a
20 microphone. The signals from the different inputs are converted, if required, into digital signals. Then, the sound card management program assigns each input 11, 12, 13, by means of a first processing circuit 111, 121, 131, with a gain 21, 22, 23 corresponding to that stored in
25 the central processing unit database linked with the recording produced. For the microphone input 13 and the auxiliary source input 12, this is a predefined gain 22, 23, set according to the characteristics of the microphone and auxiliary source. For the input 11
30 receiving the signals corresponding to the recordings, the sound card management program collects, in its database 30 stored in the central processing unit, the

gain calculated according to the formula a for the incoming recording on the first input and a gain 31 accounting for the use of an MPEG decoder, for example. These two gains are then applied to the inputs 210, 211 of a summing circuit 21, the output of which is connected to the first processing circuit 111 linked to the input 11 of the MPEG decoder. The three signals 110, 120, 130 modified in this way are then summed and mixed by a signal summer 20, to form a single signal 100. This signal 100 is then attenuated by an attenuating electronic circuit 10 of a specified fixed value. Indeed, if the sound levels of the input signals 110, 120, 130 are all similar to the sound level selected by the user, then the sum of these signals will necessarily exceed this maximum level selected by the user, hence the need to reduce the sound level of the signal resulting from the sum of the three signals 110, 120, 130 systematically so that, in the most unfavourable case, it is not greater than the maximum level selected by the user. The signal 100 is then assigned to at least one zone, e.g. three. The term zone refers to an area equipped with at least one loud speaker 61, 62, 63 connected to the sound card by means of an amplifier 51, 52, 53. For each zone, the sound level of the signal is modified according to the maximum sound level selected by the user for each of these zones. To do this, the maximum level selected by the user for each zone is previously stored in memory, for example, in a database of the central processing unit, and then, during reproduction, collected by the sound card management program and sent to an attenuating circuit 41, 42, 43 linked with each zone. Then, the signal 410 modified in this way according to the setting

of each zone may be modified again by a dynamic gain 411, 421, 431 assigned to low-pitched and/or high-pitched sounds, as described above. To do this, the sound card management program assigns a gain to the low-pitched and/or high-pitched sound signals contained in the output signal of each zone. This gain corresponds to the attenuation applied to the output signal of each zone. In other words, if the output signal of a zone is attenuated, for example by 6 dB, so as not to exceed the sound level selected by the user, the low-pitched and/or high-pitched sound signals will be increased by 6 dB. The attenuation assigned to each zone is collected by the sound card management program in a database 32 or a specific file stored in the central processing unit.

Once the dynamic low-pitched and/or high-pitched sound adjustment has been carried out, the digital signal 4110 is applied to the input of a digital/analogue converter 412, 422, 423, the output of which is connected to the input of an amplifier 51, 52, 53 on which loud speakers 61, 62, 63 are connected.

It is understood that the process according to the invention makes it possible, after prior determination of the possible volume gain for each recording, to reproduce all the digital recordings analysed, with the same sound level, for the same sound setting selected by a user.

It must be clear for those experienced in the art that the present invention enables embodiments in many other specific forms without leaving the field of the invention as claimed. Consequently, the present embodiments must be considered as illustrations, but may be modified in the field defined by the scope of the

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100